Data Collection and "Data Suitability" First Steps
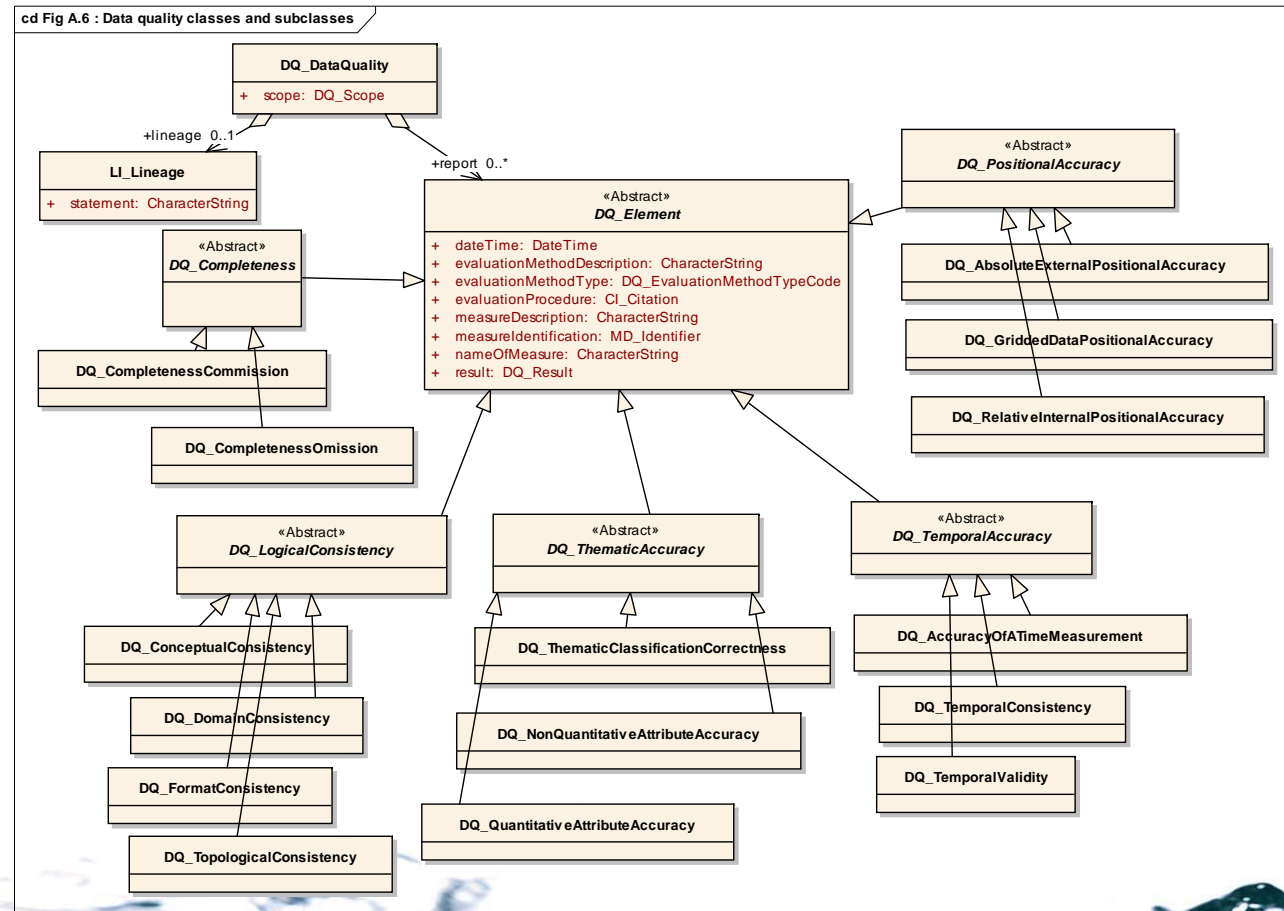
Keiran Millard

(k.millard@hrwallingford.co.uk)

HR Wallingford
Working with water

- Suitability = Quality = Fitness for purpose

- To allow users to assess up front whether data associated with the programme, activities and series are suitable for their intended use.

- What information enables users to assess whether the data is suitable for their use?

  - Primary users?
  - Secondary users?

- When primary users commission data capture, the organisation collecting the data follows procedures
    - so that the data collected meets the users needs
- Identified a range of example procedures that were used
    - Some more strongly documented than others
    - Some more widely used than others
- If secondary users knew what procedures had been followed, it would provide them with the basis to determine data suitability and hence re-use
    - But what information needs to be supplied about the procedures?

HR Wallingford
Working with water

- How show quality information be supplied?

cd Fig A.6 : Data quality classes and subclasses

- **Information required:**
  - That a defined procedure was used to collect the data
  - The name and 'citation' of the procedure
    - Web page, paper, document. Not necessarily public
  - A statement on whether the data collection complied with the procedure.

- **Information not required:**
  - Full details of the implementation of the procedure
    - Although it would be expected that organisation would retain that as part of their normal internal quality procedures

- **EOF success criteria would be that metadata contained the above quality information**

- Test EOF conclusions about quality information requirements for metadata

- To get agreement and feedback on the approach
  - Data collector perspective
  - Data user perspective

- Government understands that climate change is having a serious impact on the counter population of the UK.

- They have therefore embarked on an ambitious three-year programme to monitor and record counter populations.

- Teams will be appointed to undertake assessment of the populations in <u>four</u> different regions over the three years

HR Wallingford
Working with water

- Set up four teams
- Each year, your team will be awarded a contract to monitor the counter populations for a UK region (National Park)
  - You will not be successful in winning the same region each year!
- Each team will have
  - A survey sheet to complete
  - Data collection instructions
  - Population of counters
- Determine the health of the population of counters
  - by counting them according to your procedure
  - Record the species count on the survey sheet

Park	2007	2008	2009
Caringorm	D	C	D
New Forest	A	B	A
Peak District	B	D	C
Brecon Beacons	C	A	B

Park	2007	2008	2009
Caringorm	D **40**	C **30**	D **40**
New Forest	A **40**	B **55**	A **75**
Peak District	B **55**	D **35**	C **75**
Brecon Beacons	C **42**	A **55**	B **50**

Cairngorm Counter Masacre

- In 2009 the counter population in the Cairngorms plummeted to zero.

- Whilst populations in all other parts of the UK have remain stable, locals are worried about the situation in the Cairngorms.

- Government agencies are investigating what happened based on data collected as part of the Counters in Crisis study

Park	2007	2008	2009
Caringorm	D **40** 1.0d	C **30** ??	D **40** 1.0d
New Forest	A **40** 3.0a	B **55** 3.0a	A **75** 3.0a
Peak District	B **55** 3.0a	D **35** 1.0d	C **75** ??
Brecon Beacons	C **42** ??	A **55** 3.0a	B **50** 3.0a

Park	2007	2008	2009
Caringorm	30	25	0
New Forest	40	55	75
Peak District	55	55	75
Brecon Beacons	6	55	50

Conclusions

- **Metadata Required**
  - That a defined procedure was used to collect the data
    - Mandatory element
  - The name and 'citation' of the procedure
    - Web page, paper, document. Not necessarily public
  - A statement on whether the data collection complied with the procedure.
    - True / false (mandatory)
    - text explanation (optional)

- **Encoding quality information in metadata**
  - Valid for ISO 19115 / 19139
  - Consistent with INSPIRE metadata profile
  - Compatible with MEDIN metadata profile

- EOF success criteria would be that metadata contained the above quality information
  - Likely to be zero now…
  - Enable a registry of procedures used by different organisations and for different purposes
- Ensuring this information exists makes data quality more explicit
  - Not onerous on the data collector
  - Promotes data reuse